Bias-Variance

Nipun Batra and teaching staff

July 21, 2025

IIT Gandhinagar

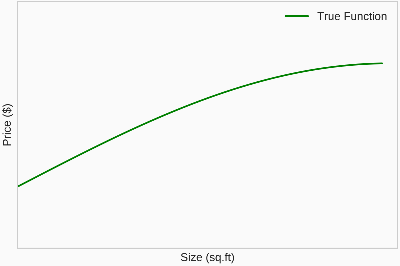
The Scenario: True Function

For the purpose of this lecture we assume that there exists a relation between **Housing Prices** and **area of the house**.

The Scenario: True Function

For the purpose of this lecture we assume that there exists a relation between **Housing Prices** and **area of the house**.

Here, the true function $f_{\boldsymbol{\theta}_{\text{true}}}$ is used to model the relation $y_t = f_{\boldsymbol{\theta}_{\text{true}}}(x_t)$



Modeling the relation

The 3 Sources of Error

Any prediction made is affected by 3 sources of error:

- Noise

The 3 Sources of Error

Any prediction made is affected by 3 sources of error:

- Noise
- Bias

The 3 Sources of Error

Any prediction made is affected by 3 sources of error:

- Noise
- Bias
- Variance

Noise

A relation between **price** and **size** will be affected by other factors that we have not considered or cannot be perfectly captured. Such factors would include:

- the condition of the house (cannot be measured perfectly)

Noise

A relation between **price** and **size** will be affected by other factors that we have not considered or cannot be perfectly captured. Such factors would include:

- the condition of the house (cannot be measured perfectly)
- sale prices of other houses in the neighborhood (measurements that have biases in themselves)

Because of this, data is inherently noisy.

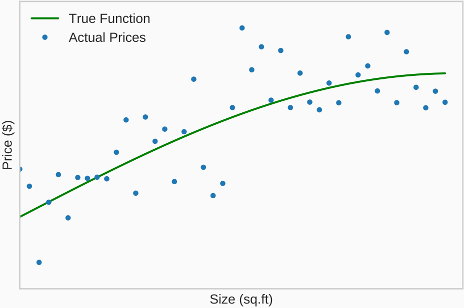
Noise

This is **not** a property of data but rather an **irreducible error**.

Noise

This is **not** a property of data but rather an **irreducible error**.

This error can be captured by the error term ϵ which causes the final value of the house to follow the equation: $y_t = f_{\theta_{\text{true}}}(x_t) + \epsilon_t$

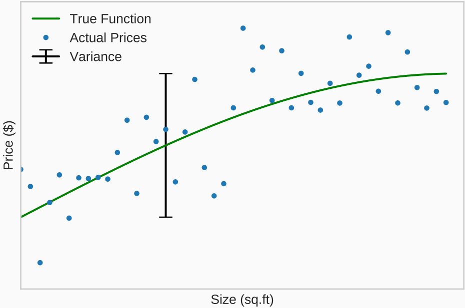


Modeling the relation

Noise

This noise can be assumed to be mean-centered around 0 with spread called the variance of the noise.

This causes y_t to become mean centered around the true relation.



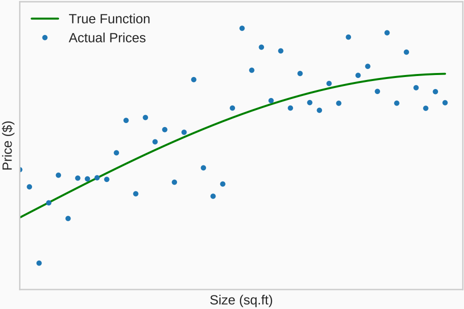
Modeling the relation

Bias

Bias is a measure of how well a model can fit a given relation.

Bias is a measure of how well a model can fit a given relation.

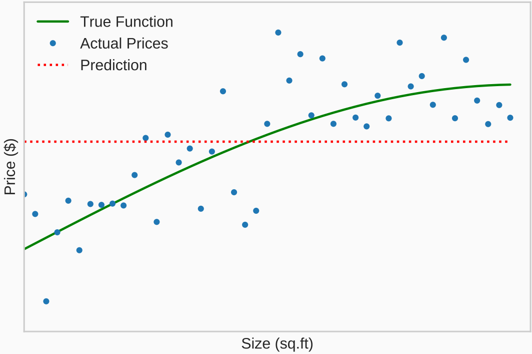
To understand this, let us take an example where we try to learn the relation that models the *Price* and *Size* of a house using a *constant function*.



An example dataset

Bias

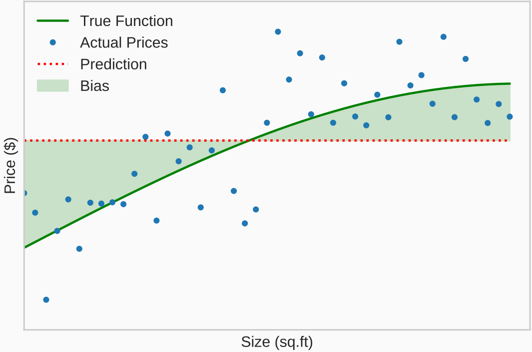
To understand this, let us take an example where we try to learn the relation that models the *Price* and *Size* of a house using a *constant function*.



An example dataset

Bias

So the bias in this scenario looks something like this:



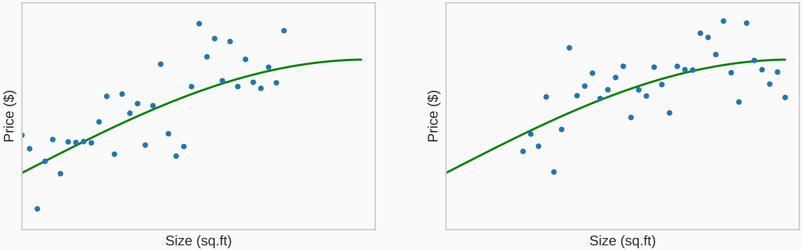
An example dataset

Bias

But it is important to understand that there are a large number of different datasets possible for a given situation, with each having their individual fits.

But it is important to understand that there are a large number of different datasets possible for a given situation, with each having their individual fits.

Assume that we have two datasets of houses sold.

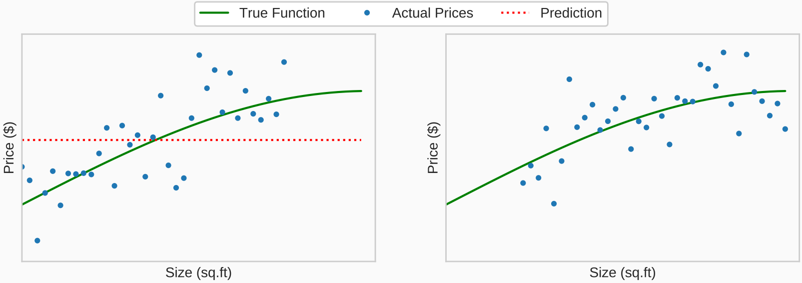


Two Datasets from same relation

Bias

But it is important to understand that there are a large number of different datasets possible for a given situation, with each having their individual fits.

If we try to fit a constant function to them.

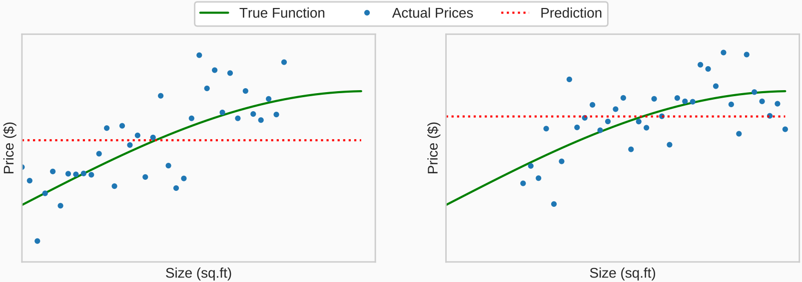


Two Datasets from same relation

Bias

But it is important to understand that there are a large number of different datasets possible for a given situation, with each having their individual fits.

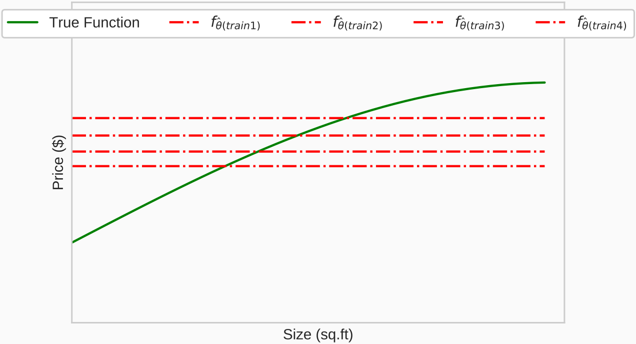
We see that they show different predictions.



Two Datasets from same relation

Bias

Doing so for all possible size N training sets we get

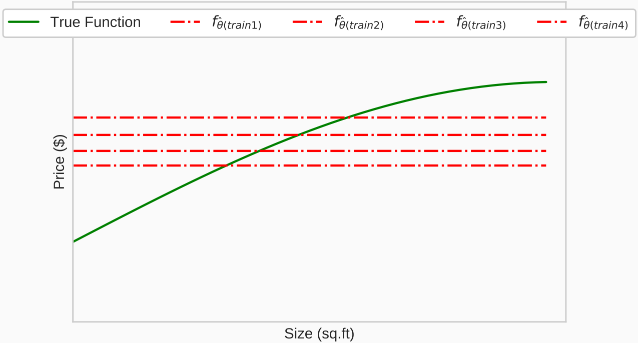


Fits on different training sets

Bias

Doing so for all possible size N training sets we get

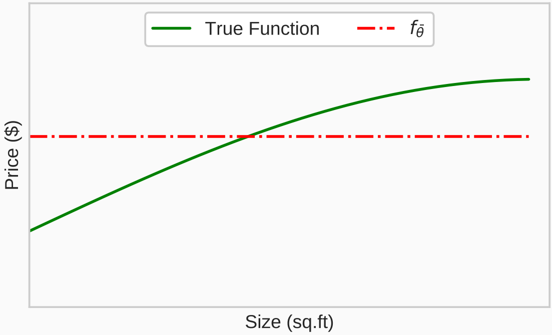
A way of consolidating all these possible fits is to calculate an average fit that is weighted by how likely they are to appear.



Fits on different training sets

Bias

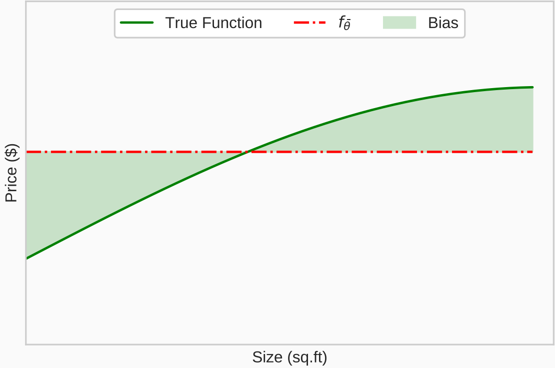
Averaging all the fits (as in this scenario all datasets are equally likely) we get the average fit.



Average fit on all different training sets

Bias Contribution

$$\text{Bias}(x) = f_{\boldsymbol{\theta}_{\text{true}}}(x) - f_{\bar{\boldsymbol{\theta}}}(x)$$

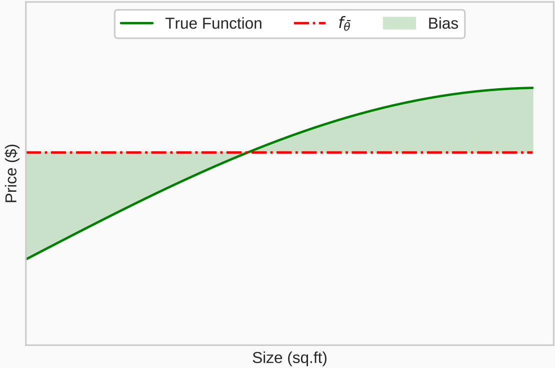


Bias of the fit

Bias Contribution

$$\text{Bias}(x) = f_{\boldsymbol{\theta}_{\text{true}}}(x) - f_{\bar{\boldsymbol{\theta}}}(x)$$

It is a measure of how flexible the fit is in capturing $f_{\boldsymbol{\theta}_{\text{true}}}(x)$



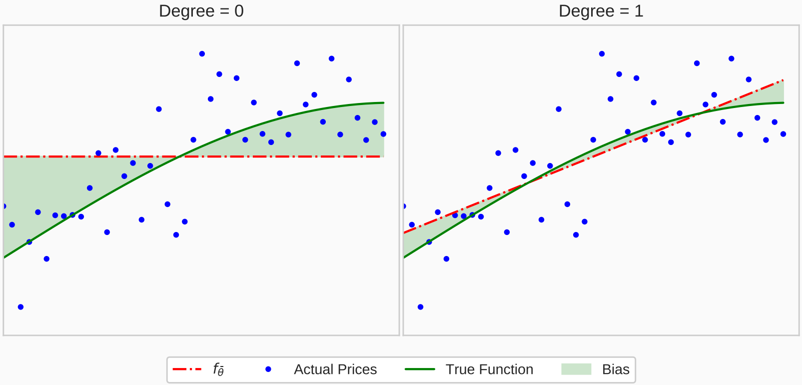
Bias of the fit

Bias Contribution: Effect of Complexity

As we increase the complexity of the fit

$\implies$ fit becomes more flexible

$\implies$ bias decreases



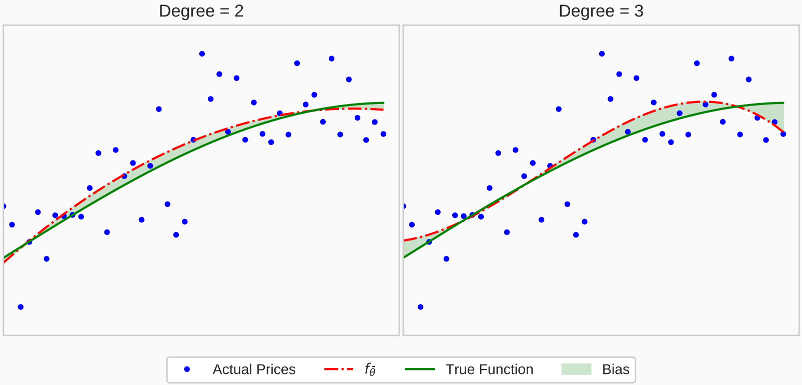
Effect of degree on Bias of the fit

Bias Contribution: Effect of Complexity

As we increase the complexity of the fit

$\implies$ fit becomes more flexible

$\implies$ bias decreases



Effect of degree on Bias of the fit

Bias: Calculating the Bias

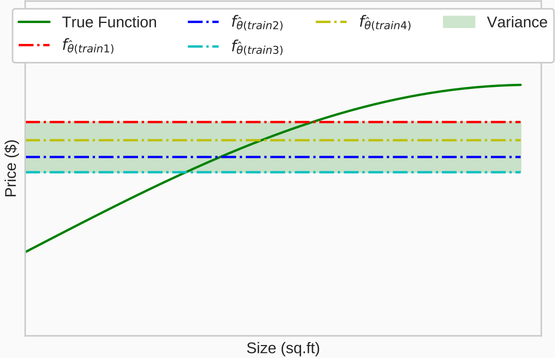
Bias calculation for a model is at the core a calculation of the area under a curve.

Therefore, finding the bias for a model in the range (a, b) is the calculation of the integral:

$$\int_a^b |f_{\bar{\theta}}(x) - f_{\theta(\text{true})}(x)| dx$$

Variance

Variance of the fit is a measure of the variation in the fits when trained across different training sets.



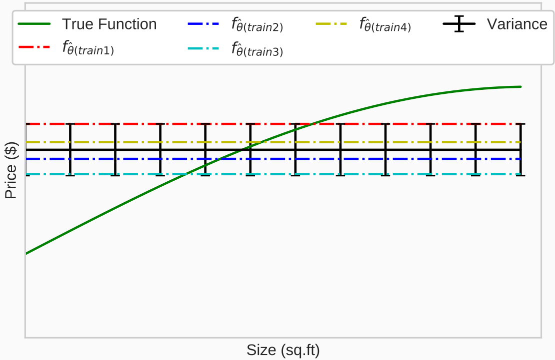
Variance of the fit

Variance Contribution

For Low Complexity

$\Longrightarrow$ variations between curves are less

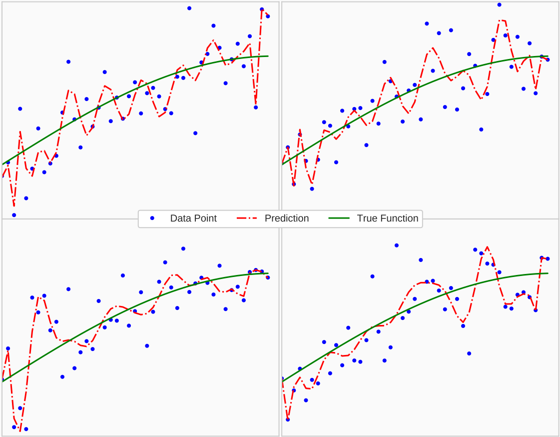
$\Longrightarrow$ Variance is less



Variance of the low complexity fits

Variance Contribution

For High Complexity we see very high variation



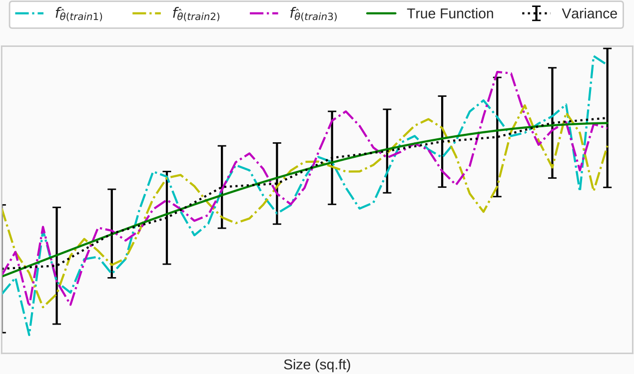
Variance in high complexity fits

Variance Contribution

For High Complexity

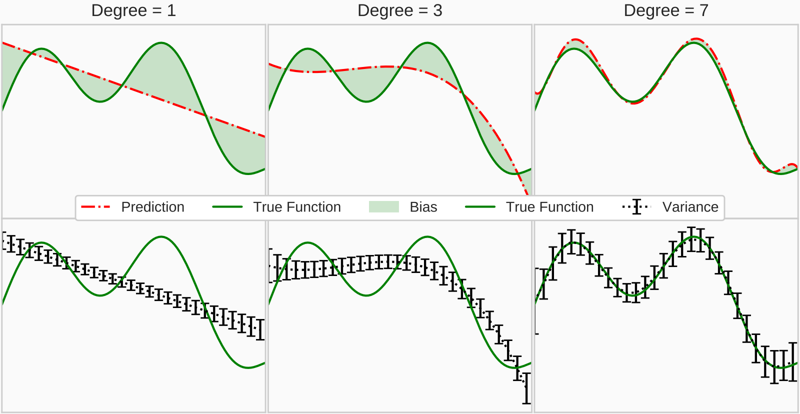
$\implies$ high variation

$\implies$ Variance is high



Variance of the high complexity fits

The Bias-Variance Trade off



Variance in high complexity fits

The Bias-Variance Trade off

Plot Graph - 3:06 Variance and the bias-variance trade off

Mathematically Formulating the Error of a Model

Measuring the goodness of a Model

To measure the goodness of a model, we have to understand how well it can predict the behavior of the phenomenon it is trying to model.

Measuring the goodness of a Model

To measure the goodness of a model, we have to understand how well it can predict the behavior of the phenomenon it is trying to model.

This behavior varies due to training set randomness.

Measuring the goodness of a Model

To measure the goodness of a model, we have to understand how well it can predict the behavior of the phenomenon it is trying to model.

This behavior varies due to training set randomness.

Therefore, it is important to measure performance **averaged over all possible training sets** (of size N).

Measuring the goodness of a Model

To measure the goodness of a model, we have to understand how well it can predict the behavior of the phenomenon it is trying to model.

This behavior varies due to training set randomness.

Therefore, it is important to measure performance **averaged over all possible training sets** (of size N).

$$E_{\text{training set}}[\text{error of } \hat{\theta}(\text{training set})]$$

gives a measure of the average error by doing an expectation of the errors of all possible training sets of size N.

Expected Prediction Error at a point

Any prediction made is affected by 3 sources of error:

- Noise

Expected Prediction Error at a point

Any prediction made is affected by 3 sources of error:

- Noise
- Bias

Expected Prediction Error at a point

Any prediction made is affected by 3 sources of error:

- Noise
- Bias
- Variance

Expected Prediction Error at a point

Any prediction made is affected by 3 sources of error:

- Noise
- Bias
- Variance

Expected Prediction Error at a point

Any prediction made is affected by 3 sources of error:

- Noise
- Bias
- Variance

Therefore, $E_{train}[\text{at a point } x_t] = f(\text{noise, bias, variance})$

Formally defining the 3 sources of error: Noise

Noise is an **irreducible error** captured by the error term ϵ.

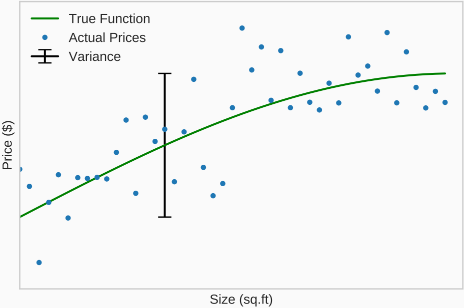
The equation of the relation becomes $y_t = f_{\theta(true)}(x_t) + \epsilon_t$

Formally defining the 3 sources of error: Noise

Noise is an **irreducible error** captured by the error term ϵ.

The equation of the relation becomes $y_t = f_{\theta(true)}(x_t) + \epsilon_t$

The noise is mean-centered around 0 with spread called the variance of the noise, denoted by σ^2.



Variance in the noise

Formally defining the 3 sources of error: Noise

Noise is an **irreducible error** captured by the error term ϵ.

The equation of the relation becomes $y_t = f_{\theta(true)}(x_t) + \epsilon_t$

The noise is mean-centered around 0 with spread called the variance of the noise, denoted by σ^2.

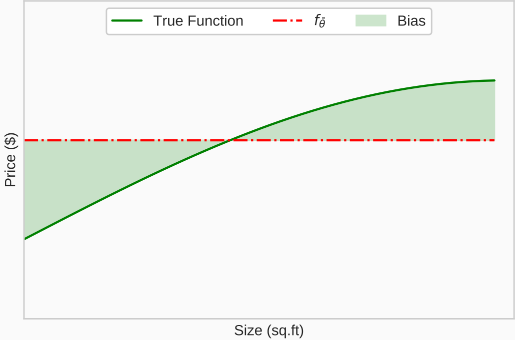
That is, it can be denoted by $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$

Formally defining the 3 sources of error: Bias

Bias is a measure of how flexible the fit is in capturing the true function $f_{\boldsymbol{\theta}_{\text{true}}}(x)$

$$\text{Bias}(x_t) = f_{\boldsymbol{\theta}_{\text{true}}}(x_t) - f_{\bar{\boldsymbol{\theta}}}(x_t)$$

where $f_{\bar{\boldsymbol{\theta}}}$ denotes the average fit over all datasets.



Bias of the fit

Formally defining the 3 sources of error: Bias

Bias is a measure of how flexible the fit is in capturing the true function $f_{\boldsymbol{\theta}_{\text{true}}}(x)$

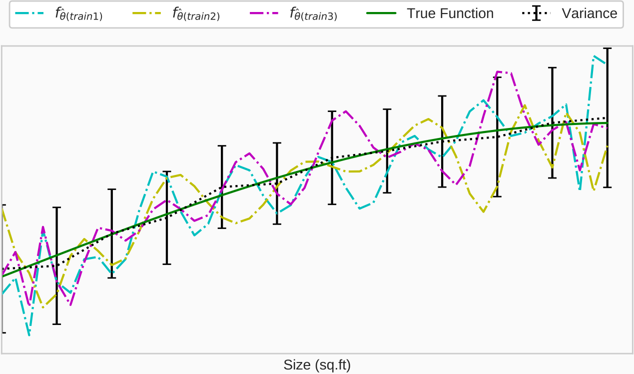
$$\text{Bias}(x_t) = f_{\boldsymbol{\theta}_{\text{true}}}(x_t) - f_{\bar{\boldsymbol{\theta}}}(x_t)$$

where $f_{\bar{\boldsymbol{\theta}}}$ denotes the average fit over all datasets.

As $f_{\bar{\boldsymbol{\theta}}}$ denotes the average fit over all datasets, it can be expressed by $f_{\bar{\boldsymbol{\theta}}}(x_t) = E_{\text{train}}[f_{\hat{\boldsymbol{\theta}}}(x_t)]$

Formally defining the 3 sources of error: Variance

Variance of the fit is a measure of the variation in the fits when trained across different training sets.



Variance of a fit

Formally defining the 3 sources of error: Variance

Variance of the fit is a measure of the variation in the fits when trained across different training sets.

Variance of the fit can be defined by

$$\mathrm{var}(f_{\hat{\boldsymbol{\theta}}}(x_t)) = E_{\text{train}}[(f_{\hat{\boldsymbol{\theta}}}(x) - f_{\bar{\boldsymbol{\theta}}}(x_t))^2]$$

where $f_{\hat{\boldsymbol{\theta}}}(x) - f_{\bar{\boldsymbol{\theta}}}(x_t)$ denotes the deviation that a specific fit has from the average.

Deriving Expected Prediction Error

Now we will see how,

$E_{train}[\text{at a point } x_t] = \sigma^2 + [\text{bias}(f_{\hat{\theta}}(x_t))]^2 + \text{var}(f_{\hat{\theta}}(x_t))$

where,

given a training set, the parameters $\hat{\theta}$ of the fit are learned as $f_{\hat{\theta}}$

and, the prediction at a point x_t for the model trained on that training set is $f_{\hat{\theta}}(x_t)$

Deriving Expected Prediction Error

Prediction Error at a point x_t can be calculated using the squared loss function.

Prediction error at $x_t = (y_t - f_{\hat{\theta}(train)}(x_t))^2$

To find the "Expected Prediction Error" at a point x_t we average out the prediction error at that point over all possible learned models. This can be done by finding the expectation of prediction error for that point over all possible training datasets (*train*) and labels for that point (y_t).

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train, y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$= E_{train, y_t}[((y_t - f_{\theta(true)}(x_t)) + (f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)))^2]$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[(\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\text{a}} + \underbrace{(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\text{b}})^2]$$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[(\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\text{a}} + \underbrace{(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)))}_{\text{b}})^2]$$

$$= E_{train,y_t}[(a+b)^2]$$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[(\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\text{a}} + \underbrace{(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\text{b}})^2]$$

$$= E_{train,y_t}[(a+b)^2]$$

$$= E_{train,y_t}[a^2 + 2ab + b^2]$$

Deriving Expected Prediction Error

Expected prediction error at $x_t = E_{train,y_t}[(y_t - f_{\hat{\theta}(train)}(x_t))^2]$

$$= E_{train,y_t}[(\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\text{a}} + \underbrace{(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\text{b}})^2]$$

$$= E_{train,y_t}[(a+b)^2]$$

$$= E_{train,y_t}[a^2 + 2ab + b^2]$$

(Using Linearity of Expectation)
$= E_{train,y_t}[a^2] + 2E_{train,y_t}[ab] + E_{train,y_t}[b^2]$.......................(Eqn. 1)

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

(Since there is no dependence on training set)

$$= E_{y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

$$(\because \text{ there is no dependence on training set})$$

$$= E_{y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))^2}_{\epsilon_t^2}]$$

$$= E_{y_t}[\epsilon_t^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

$$(\because \text{ there is no dependence on training set})$$

$$= E_{y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))^2}_{\epsilon_t^2}]$$

$$= E_{y_t}[\epsilon_t^2]$$

$$= \sigma^2 (\text{By definition})$$

Deriving Expected Prediction Error

$$E_{train,y_t}[a^2] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))^2]$$

$$(\because \text{ there is no dependence on training set})$$

$$= E_{y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))^2}_{\epsilon_t^2}]$$

$$= E_{y_t}[\epsilon_t^2]$$

$$= \sigma^2 (\text{By definition})$$

$$E_{train,y_t}[a^2] = \sigma^2 \ldots\ldots\ldots\ldots\ldots\ldots (\text{Eqn. } 2)$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[(y_t - f_{\theta(true)}(x_t))(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train,y_t}[\epsilon_t(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train,y_t}[\epsilon_t(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$(\because \epsilon_t \text{ and } (f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)) \text{ are independent})$$

$$= E_{train,y_t}[\epsilon_t] \times E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train,y_t}[\epsilon_t(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$(\because \epsilon_t$ and $(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))$ are independent)

$$= \underbrace{E_{train,y_t}[\epsilon_t]}_{=0} \times E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

(By definition ϵ_t has mean 0)

Deriving Expected Prediction Error

$$E_{train,y_t}[ab] = E_{train,y_t}[\underbrace{(y_t - f_{\theta(true)}(x_t))}_{\epsilon_t}(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train,y_t}[\epsilon_t(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

($\because \epsilon_t$ and $(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))$ are independent)

$$= \underbrace{E_{train,y_t}[\epsilon_t]}_{=0} \times E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

(By definition ϵ_t has mean 0)

$E_{train,y_t}[ab] = 0$.............(Eqn. 3)

Deriving Expected Prediction Error

$$E_{train,y_t}[b^2] = E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[b^2] = E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)$ is independent of $y_t)$

$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train,y_t}[b^2] = E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)$ is independent of y_t)

$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= MSE(f_{\hat{\theta}(train)}(x_t))$$

Deriving Expected Prediction Error

$$E_{train,y_t}[b^2] = E_{train,y_t}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t)$ is independent of $y_t)$

$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= MSE(f_{\hat{\theta}(train)}(x_t))$$

$E_{train,y_t}[b^2] = MSE(f_{\hat{\theta}(train)}(x_t))$ .... (Eqn. 4)

Deriving Expected Prediction Error

From Eqn. 1, 2, 3 and 4, we get,

Expected prediction error at $x_t = \sigma^2 + MSE(f_{\hat{\theta}(train)}(x_t))$

Now, we will further simplify the MSE term into bias and variance.

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[((f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t)) + (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)))^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(\underbrace{(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))}_{\alpha} + \underbrace{(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\beta})^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(\underbrace{(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))}_{\alpha} + \underbrace{(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\beta})^2]$$

$$= E_{train}[(\alpha + \beta)^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(\underbrace{(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))}_{\alpha} + \underbrace{(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\beta})^2]$$

$$= E_{train}[(\alpha + \beta)^2]$$

$$= E_{train}[\alpha^2 + 2\alpha\beta + \beta^2]$$

Deriving Expected Prediction Error

$$MSE(f_{\hat{\theta}(train)}(x_t)) = E_{train}[(f_{\theta(true)}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(\underbrace{(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))}_{\alpha} + \underbrace{(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))}_{\beta})^2]$$

$$= E_{train}[(\alpha + \beta)^2]$$

$$= E_{train}[\alpha^2 + 2\alpha\beta + \beta^2]$$

(Using Linearity of Expectation)

$= E_{train}[\alpha^2] + 2E_{train}[\alpha\beta] + E_{train}[\beta^2]$ ....(Eqn. 5)

Deriving Expected Prediction Error

$$E_{train}[\alpha^2] = E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train}[\alpha^2] = E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\theta(true)}(x_t) - E_{train}[f_{\hat{\theta}(train)}(x_t)]^2]$$

Deriving Expected Prediction Error

$$
\begin{aligned}
E_{train}[\alpha^2] &= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2] \\
&= E_{train}[(f_{\theta(true)}(x_t) - E_{train}[f_{\hat{\theta}(train)}(x_t)]^2] \\
&= E_{train}[bias(f_{\hat{\theta}}(x_t))^2] && \text{(By definition of bias)}
\end{aligned}
$$

Deriving Expected Prediction Error

$$E_{train}[\alpha^2] = E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\theta(true)}(x_t) - E_{train}[f_{\hat{\theta}(train)}(x_t)]^2]$$

$$= E_{train}[bias(f_{\hat{\theta}}(x_t))^2] \qquad \text{(By definition of bias)}$$

$$= bias(f_{\hat{\theta}}(x_t))^2$$

($\because$ bias is not a function of training data)

Deriving Expected Prediction Error

$$E_{train}[\alpha^2] = E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\theta(true)}(x_t) - E_{train}[f_{\hat{\theta}(train)}(x_t)]^2]$$

$$= E_{train}[bias(f_{\hat{\theta}}(x_t))^2] \qquad \text{(By definition of bias)}$$

$$= bias(f_{\hat{\theta}}(x_t))^2$$

($\because$ bias is not a function of training data)

$$E_{train}[\alpha^2] = bias(f_{\hat{\theta}}(x_t))^2 \text{ ....(Eqn. 6)}$$

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= bias_t \times E_{train}[f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)]$$

($\because$ bias$_t$ is not a function of training data)

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= bias_t \times E_{train}[f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)]$$

($\because$ bias_t is not a function of training data)

$$= bias \times \left(E_{train}[f_{\bar{\theta}}(x_t)] - E_{train}[f_{\hat{\theta}(train)}(x_t)]\right)$$

Deriving Expected Prediction Error

$$E_{train}[\alpha\beta]$$
$$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$$

$$= bias_t \times E_{train}[f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)]$$

($\because$ bias_t is not a function of training data)

$$= bias \times \left(E_{train}[f_{\bar{\theta}}(x_t)] - E_{train}[f_{\hat{\theta}(train)}(x_t)]\right)$$

$$= bias \times (f_{\bar{\theta}}(x_t) - f_{\bar{\theta}}(x_t))$$

($\because$ $f_{\bar{\theta}}(x_t) = E_{train}[f_{\hat{\theta}(train)}(x_t)$)

Deriving Expected Prediction Error

$E_{train}[\alpha\beta]$
$= E_{train}[(f_{\theta(true)}(x_t) - f_{\bar{\theta}}(x_t))(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$

$= E_{train}[bias_t \times (f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))]$

$= bias_t \times E_{train}[f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t)]$

($\because$ bias$_t$ is not a function of training data)

$= bias \times \left(E_{train}[f_{\bar{\theta}}(x_t)] - E_{train}[f_{\hat{\theta}(train)}(x_t)]\right)$

$= bias \times (f_{\bar{\theta}}(x_t) - f_{\bar{\theta}}(x_t))$

$E_{train}[\alpha\beta] = 0$..........................(Eqn. 7)

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - E_{train}[(f_{\hat{\theta}(train)}(x_t)])^2]$$

$$(\because f_{\bar{\theta}}(x_t) = E_{train}[(f_{\hat{\theta}(train)}(x_t)]\ )$$

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - E_{train}[(f_{\hat{\theta}(train)}(x_t)])^2]$$

$$(\because f_{\bar{\theta}}(x_t) = E_{train}[(f_{\hat{\theta}(train)}(x_t)]\ )$$

$$= variance(f_{\hat{\theta}}(x_t))$$

Deriving Expected Prediction Error

$$E_{train}[\beta^2] = E_{train}[(f_{\bar{\theta}}(x_t) - f_{\hat{\theta}(train)}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - f_{\bar{\theta}}(x_t))^2]$$

$$= E_{train}[(f_{\hat{\theta}(train)}(x_t) - E_{train}[(f_{\hat{\theta}(train)}(x_t)])^2]$$

$$(\because f_{\bar{\theta}}(x_t) = E_{train}[(f_{\hat{\theta}(train)}(x_t)]\ )$$

$$= variance(f_{\hat{\theta}}(x_t))$$

$$E_{train}[\beta^2] = variance(f_{\hat{\theta}}(x_t))\text{...............(Eqn. 8)}$$

Deriving Expected Prediction Error

From Eqn. 1 - 8, we get,

Expected prediction error at x_t

$$= \sigma^2 + MSE(f_{\hat{\theta}(train)}(x_t))$$

$$= \sigma^2 + bias(f_{\hat{\theta}}(x_t))^2 + variance(f_{\hat{\theta}}(x_t))$$